Fast Clustering in Linear 1D Subspaces: Segmentation of Microscopic Image of Unstained Specimens

Ivica Kopriva[1], Maria Brbić[1], Dijana Tolić[1], Nino Antulov-Fantulin[1] and Xinjian Chen[2]
[1]Division of Electronics, Ruđer Bošković Institute, Bijenička cesta 54, P.O. Box 180, 10002, Zagreb, Croatia
[2]School of Electronic and Information Engineering, Soochow University, Suzhou 215123, China

e-mail: ikopriva@irb.hr, maria.brbic@irb.hr, Dijana.Tolic@irb.hr, Nino.Antulov.Fantulin@irb.hr, xjchen@suda.edu.cn

Abstract

Algorithms for subspace clustering (SC) are effective in terms of the accuracy but exhibit high computational complexity. We propose algorithm for SC of (highly) similar data points drawn from union of linear one-dimensional subspaces that are possibly dependent in the input data space. The algorithm finds a dictionary that represents data in reproducible kernel Hilbert space (RKHS). Afterwards, data are projected into RKHS by using empirical kernel map (EKM). Due to dimensionality expansion effect of the EKM one-dimensional subspaces become independent in RKHS. Segmentation into subspaces is realized by applying the *max* operator on projected data which yields the computational complexity of the algorithm that is linear in number of data points. We prove that for noise free data proposed approach yields exact clustering into subspaces. We also prove that EKM-based projection yields less correlated data points. Due to nonlinear projection, the proposed method can adopt to linearly nonseparable data points. We demonstrate accuracy and computational efficiency of the proposed algorithm on synthetic dataset as well as on segmentation of the image of unstained specimen in histopathology.

Keywords: Subspace clustering, 1D subspaces, empirical kernel map, segmentation, unstained specimen.

1 INTRODUCTION

Low-contrast images, such as color microscopic image of unstained histological specimen, are composed of objects with highly correlated spectral profiles. Segmentation of an image of unstained specimen is motivated by a practical reasons. It is an important step in design of computer aided diagnostic system based on the image of unstained histopathological specimen, whereas obtained segments are annotated by pathologist and used to train the classifier. However, images of unstained specimen are very hard to segment for many state-of-the-art image segmentation algorithms [1]. Color microscopic image of a specimen is composed of 3D pixels where each pixel is occupied by one object (tissue) only [1], i.e. the pixels are pure. Segmentation of microscopic image of histopathological specimen can be executed through the subspace clustering (SC) [2], where each subspace corresponds to a single tissue component. However, segmentation of such image is also very challenging for state-of-the-art SC methods due to two reasons: (1) algorithms for SC such as sparse SC [3, 2], and low-rank representation (LRR) SC [4,

Medical Imaging 2017: Digital Pathology, edited by Metin N. Gurcan, John E. Tomaszewski, Proc. of SPIE Vol. 10140, 1014003 · © 2017 SPIE · CCC code: 1605-7422/17/$18 · doi: 10.1117/12.2247806

2] are effective in terms of the accuracy but suffer from high computational complexity associated with construction of similarity matrix and spectral clustering [5]. For a dataset consisting of N points, they have complexities in respective order of $O(N^2)$ and $O(N^3)$. Thus, cited methods as well as SC methods [6, 7, 8, 9] are computationally intractable for segmentation of microscopic images of pathological specimens comprised of $N>10^6$ pixels; (2) due to very high similarity between the pixels in microscopic image of unstained specimen, all data points (pixels) are extremely close to the intersection of the subspaces. Thus, presence of the small amount of noise will cause incorrect assignment of data points to the subspaces. This sets the motivation to develop fast algorithm for SC of highly similar data points belonging to linear one-dimensional subspaces with computational complexity of $O(N)$.

2 MATERIALS AND METHODS

We follow problem definition given in [1]. Let $\mathbf{X}=\left[\mathbf{x}_1,...,\mathbf{x}_N\right]\in\mathbb{R}^{D\times N}$ be a collection of N data points drawn from union of $L\geq 1$ linear subspaces $\left\{S_i\right\}_{i=1}^{L}$ of dimensions $d_i=\dim\left(S_i\right)$, $0<d_i<D$, $i=1,...,L$. Thus, $\mathbf{X}=\cup_{i=1}^{L}\mathbf{X}_i$, where $\mathbf{X}_i\in\mathbb{R}^{D\times N_i}\doteq\left\{\mathbf{x}_{i(j)}\in\mathbb{R}^{D\times 1}:\mathbf{x}_{i(j)}\in S_i\right\}_{j=1}^{N_i}$, $\sum_{i=1}^{L}N_i=N$ and $rank(\mathbf{X}_i)=d_i$. The linear subspaces are described with:

$$S_i=\left\{\mathbf{x}_i\in\mathbb{R}^{D\times 1}:\mathbf{x}=\mu_i+\mathbf{A}_i\mathbf{z}\right\},\ i=1,...,L. \tag{1}$$

The goal of subspace clustering (SC) is to find the number of subspaces L, their dimensions $\left\{d_i\right\}_{i=1}^{L}$, the subspace bases $\left\{\mathbf{A}_i\in\mathbb{R}^{D\times d_i}\right\}_{i=1}^{L}$, and the segmentation of the points according to the subspaces. Because an affine subspace of dimension d in $\mathbb{R}^D$ can be interpreted as a linear subspace of dimension $d+1$ in $\mathbb{R}^{D+1}$ (the homogenous coordinates of $\mathbf{x}\in\mathbb{R}^D$ are given by $\left[\mathbf{x}^T\ 1\right]^T\in\mathbb{R}^{D+1}$) we shall limit our exposition to the linear subspaces only. L subspaces are called independent if and only if $\dim\left(\oplus_{i=1}^{L}S_i\right)=\sum_{i=1}^{L}d_i$. $\left\{S_i\right\}_{i=1}^{L}$ are said to be disjoint if every two subspaces intersect only at the origin. For disjoint linear subspaces it applies $\dim\left(\oplus_{i=1}^{L}S_i\right)\leq\sum_{i=1}^{L}d_i$. Hence disjointness is a more general assumption for a subspace set. If the linear subspaces are independent it applies: $rank\left(\mathbf{X}\right)=\sum_{i=1}^{L}d_i\leq\min(D,N)$ while for disjoint subspaces it applies $rank\left(\mathbf{X}\right)\leq\sum_{i=1}^{L}d_i\leq\min(D,N)$. Evidently, for $L>D$ subspaces must be dependent. Within the scope of the current paper we make the following assumptions: A1) subspaces are one-dimensional: $\left\{d_i=1\right\}_{i=1}^{L}$; A2) the number of subspaces L is known; A3) without loss of generality we assume that data $\mathbf{X}$ are nonnegative satisfying: $\left\{0\leq x_{dn}\leq 1\right\}_{d,n=1}^{D,N}$. Thus, dataset $\mathbf{X}$ has representation:

$$\mathbf{X} = \mathbf{AZ} \tag{2}$$

where $\mathbf{A} \in \mathbb{R}_{0+}^{D\times L}$ represents basis and $\mathbf{Z} \in \mathbb{R}_{0+}^{L\times N}$ is a low-dimensional representation in basis **A**. Let $\kappa(\mathbf{x},\mathbf{y})$ denote symmetric positive semi-definite kernel function. According to [11], $\kappa(\mathbf{x},\mathbf{y})$ induces a mapping $\phi:\mathbb{R}^D \to H_\kappa$ in reproducible kernel Hilbert space such that for $\mathbf{x},\mathbf{y} \in \mathbb{R}^D$, we have $\kappa(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y})$. That is known as *kernel trick*. The nonlinear mapping $\phi(\mathbf{x})$ is called explicit feature map (EFM). The empirical kernel map (EKM) $\psi(\mathbf{x}_n)$ is obtained by projection of EFM $\phi(\mathbf{x}_n)$ on *L*-dimensional subspace of H_κ:

$$\psi(\mathbf{x}_n) = \left[\phi(\mathbf{a}_1) \ldots \phi(\mathbf{a}_L)\right]^T \phi(\mathbf{x}_n) = \left[\kappa(\mathbf{a}_1,\mathbf{x}_n) \ldots \kappa(\mathbf{a}_L,\mathbf{x}_n)\right]^T \quad \forall n=1,\ldots,N \tag{3}$$

Thus, EKM projects set of data points **X** according to:

$$\mathbf{X} \in \mathbb{R}_{0+}^{D\times N} \overset{EKM}{\to} \psi(\mathbf{X}) \in \mathbb{R}_{0+}^{L\times N} \tag{4}$$

where operator $\psi(\mathbf{X})$ acts on **X** column-wise in accordance with (3). Since $L, D = const \ll N$, computational complexity of the EKM projection is O(*N*).

Remark 1. While in the input space with representation (2) it is necessary for independent one-dimensional subspaces to have *L*<*D*, in representation (4) in EKM-induced space one-dimensional subspaces will be independent even if *L*≥*D*. That is because in EKM-induced space the independence condition is given with $rank\left(\Psi(\mathbf{X})\right) = \sum_{i=1}^{L} d_i \le \min(L,N)$ and that is satisfied for $\{d_i = 1\}_{i=1}^{L}$.

Lemma 1. When representation (2) is exact, the coefficient matrix **Z** satisfies: $z_{in} = \begin{cases} 1 & \text{if } \mathbf{x}_n \in S_i \\ 0 & \text{otherwise} \end{cases}$, *i*=1,..., *L* and *n*=1,..., *N*, or equivalently $z_{in} z_{jn} = \delta(i-j)$, $\forall i,j = 1,\ldots,L \;\; \forall n = 1,\ldots,N$ where δ stands for Kronecker's delta.

Proof. When representation (2) is exact the pixels belonging to the same subspace are spatially homogenous, i.e. there is no variation between them. Thus, it follows $\mathbf{x}_n = \mathbf{a}_{i_n} \; \forall n = 1,\ldots,N$, where $1 \le i_n \le L$ represents index of a subspace to which data point $\mathbf{x}_n$ belongs.

Lemma 2. Under lemma 1, the EFM projects representation **X=AZ** in (2) into

$$\phi(\mathbf{X}) = \phi(\mathbf{A})\mathbf{Z} \tag{5}$$

whereat EFM is applied on **X** column-wise. Hence, the EFM-based projection is invariant with respect to binary latent variables **Z** that satisfy (5) [1].

Proof. The proof is presented in [1, 12].

Lemma 3. Under lemma 1, the EKM projects representation **X**=**AZ** in (2) into

$$\psi(\mathbf{X}) = \psi(\mathbf{A})\mathbf{Z} \tag{6}$$

that is, the EKM-based projection is invariant with respect to binary latent variables **Z** that satisfy (5) [1].

Proof. The proof combines EKM definition in eq.(3) with the lemma 2.

$$\psi(\mathbf{X}) = \phi(\mathbf{A})^T \phi(\mathbf{X}) = \underbrace{\phi(\mathbf{A})^T \phi(\mathbf{A})}_{\psi(\mathbf{A})}\mathbf{Z} = \psi(\mathbf{A})\mathbf{Z}.$$

Before stating the lemma 4 we introduce the mutual coherence, for example of the basis matrix $\mathbf{A} \in \mathbb{R}_{0+}^{D\times L}$ as $\mu(\mathbf{A})$ [13]:

$$\mu(\mathbf{A}) = \max_{\substack{1\le i,j\le L \\ i\ne j}} \left|\mathbf{a}_i^T \mathbf{a}_j\right| \Big/ \left(\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2\right) \tag{7}$$

Evidently, $0\le\mu(\mathbf{A})\le 1$. $\mu(\mathbf{A})$ measures the maximal similarity between the vectors $\left\{\mathbf{a}_k \in \mathbb{R}_{0+}^{D}\right\}_{k=1}^{L}$. When $\mu(\mathbf{A})$ is very close to 1 many data points will sit close to the intersection of the subspaces. That is especially the problem when data are contaminated by noise and it becomes very demanding for the SC to deal with such data points. It is of great importance that EKM-based nonlinear projection of **X** can reduce correlation between data points in mapped space.

Lemma 4. When the representation (2) is exact, i.e. **X**=**AZ**, the EKM (3)/(4) with properly chosen kernel function, such as Gaussian kernel $\kappa(\mathbf{a}_k, \mathbf{x}_n) = \exp\left(-\|\mathbf{x}_n - \mathbf{a}_k\|_2^2 \big/ 2h^2\right)$, projects data **X** into $\psi(\mathbf{X})$ such that:

$$\mu(\psi(\mathbf{A})) < \mu(\mathbf{A}). \tag{8}$$

Proof. When (2) is the exact representation of data, the EKM-based projection of **X** is according to

lemma 3: $\psi(\mathbf{X}) = \psi(\mathbf{A})\mathbf{Z}$. Based on lemma 1 it follows $\psi(\mathbf{x}_n) = \psi(\mathbf{a}_{i_n})\ \forall n = 1,...,N$. From that it follows $\kappa(\mathbf{a}_k, \mathbf{x}_n) = \kappa(\mathbf{a}_k, \mathbf{a}_{i_n})\ \forall k = 1,...,L$. Since every basis vector faithfully represents the group of data points belonging to the corresponding subspace, it is possible for Gaussian kernel to select the small enough bandwidth *h* such that $\kappa(\mathbf{a}_k, \mathbf{a}_{i_n}) \to 0$ for $k \neq i_n$, where $1 \leq k, i_n \leq L$. Since for $k = i_n$: $\kappa(\mathbf{a}_{i_n}, \mathbf{a}_{i_n}) = 1$, the column vectors of Ψ(**A**) will be better separated than column vectors of **A**.

Theorem 1. SC of data points $\{\mathbf{x}_n\}_{n=1}^{N}$ belonging to linear one-dimensional subspaces is obtained by

$$i_n = \underbrace{\arg\max}_{k=1,...,L} \kappa(\mathbf{a}_k, \mathbf{x}_n)\ \ \forall n = 1,...,N \tag{9}$$

where $1 \leq i_n \leq L$ represents index of a subspace to which data point $\mathbf{x}_n$ belongs and $\kappa(\mathbf{a}_k, \mathbf{x}_n) = \psi_{kn}$. Since $L = const \ll N$, computational complexity of subspace assignment method (9) is O(*N*).

Proof. By taking into account lemma 1, lemma 2, eq. (5), and lemma 3, eq. (6), it follows:

$$\psi(\mathbf{x}_n) = \phi(\mathbf{A})^T \phi(\mathbf{A})\mathbf{z}_n = \psi(\mathbf{A})\mathbf{z}_n = \psi(\mathbf{a}_{i_n})$$

where $1 \leq i_n \leq L$ represents index of a subspace to which data point $\mathbf{x}_n$ belongs. By using the same reasoning as in the proof of lemma 4 we conclude that criterion (9) finds subspace index $1 \leq i_n \leq L$ to which data point $\mathbf{x}_n$ belongs.

Overall computational complexity of the algorithm is O(*N*). Probability of selecting dictionary atom representing object with the size of *P* pixels from $\sqrt{N}$ data points selected uniformly at random is approximately $1-(1-P/N)^{\sqrt{N}}$. Thus, for 1 megapixel image and *P/N*=0.5% selection probability is 99.33%. We name the proposed method fast EKM low-rank subspace clustering (FELRSC) algorithm. The algorithm is summarized in Algorithm 1. We demonstrate performance of proposed method on synthetic dataset as well as on segmentation of microscopic images of unstained specimen of human liver with hepatocellular carcinoma and human liver with metastasis of colon cancer. Thereby, size of the images was $N \approx 1.44 \times 10^6$ pixels. The code and data are available at [10]. Comparative analysis has been carried out in Matlab on a computer with 64GB RAM, i7-3930K CPU with 3.2GHz.

Algorithm 1. The FELRSC algorithm.
Input: Multichannel image $\mathbf{X} \in \mathbb{R}_{0+}^{D \times N}$. Number of subspaces L.
Step 1. Find dictionary **A** in (2).
Step 2. Execute EKM $\Psi(\mathbf{X}) \in \mathbb{R}_{0+}^{L \times N}$ in (3).
Step 3. Apply the *max* operator column wise to $\Psi(\mathbf{X})$ such as in (9) to obtain subspace membership index set.
Output: $i_n \in \{1,...,L\}$ $n = 1,...,N$ subspace membership index set.

3 RESULTS

We demonstrate the algorithm on synthetic dataset as well as on segmentation of color microscopic images of unstained specimens of human liver with hepatocellular carcinoma (HCC) and metastasis of colon cancer [1]. Synthetic data were, according to (2), generated from the union of L=5 one-dimensional linear subspaces with D=3 and N=10000. The dictionary **A** has been generated such that $\mu(\mathbf{A})$=0.9995. Thus, data points were highly correlated. That is illustrated in Figure 1a, whereas data are organized as RGB image with 100×100 pixels. In addition to the FELRSC algorithm, we have tested the closed form solution for LRR-SC for data contaminated by noise [14], with hand tuned threshold parameter, the nearest subspace neighbor (NSN) greedy subspace recovery (GSR) SC algorithm [7] with a MATLAB code downloaded from [15], as well as the landmark based SC (LSC) algorithm [16] with a MATLAB code downloaded from [17]. The Gaussian kernel-based nonlinear EKM was used for data projection. We present results, in terms of accuracy averaged over 10 runs, for various SNRs in Table 1. To make comparison fair, we have hand tuned number of neighbors vs. SNR value for NSN-GSR and LSC algorithms. Note, however, that it becomes highly demanding to select these values on experimental data. We point out that $\mu(\Psi(\mathbf{A}))\approx 0.9807$, which confirms capability of the EKM projection to decrease similarity between data points. That explains improved accuracy of the LRR and NSN-GSR algorithms when applied on EKM-mapped data and shows that EKM projection can be used to improve other algorithms. Computation times in the order corresponding with the table 1 were for SNR=20dB: 3.1±0.09s, 1.1±0.1s, 57.3±1.35s, 60.3±1.1s, 9.67±0.16s, 13.13±0.22s. The dictionary **A** was estimated by feature vectors selection (FVS) method [18] from all data points. Computation times for other SNR values were very close to the presented ones. Thus, in addition to the highest accuracy, the proposed algorithm also exhibits competitive computational efficiency. We have applied the proposed algorithm and the LSC algorithm to segmentation of microscopic image of unstained specimen of human liver with HCC, Fig. 2a. The image is comprised of N=1037×1388≈1.44×10^6 pixels. Thus, spectral clustering-based SC algorithms such as [7][8][10][11][14] are computationally intractable in this scenario. We have estimated from Fig. 2a that $\mu(\mathbf{A})$>0.9999 [1]. Thus, pixels are highly correlated and that makes them difficult to segment. The main tissues presented in the image are HCC, tumor fibrotic capsule and blood vessel [1]. The ground truth image related to image of unstained specimen of HCC is obtained by immunohistochemical staining for hepatocyte antigen: Hepatocyte Clone OCH1E5 (Hep Par), of the subsequent slide, see [1] for in-depth description, and it is shown in Fig. 2b. Therein, HCC (hepatocytes) is colored brown, endothelium of blood vessel is colored blue, and tumor fibrotic capsule is colored white.

The color coded result obtained in [1] by applying the ℓ_0-norm constrained NMF algorithm on EKM projected image, the EKM-NMF_L0 algorithm, is shown in Fig. 2c, whereat HCC tissue is colored in blue, blood vessel in red and tumor fibrotic capsule in green. Color coded results obtained by the FELRSC algorithm is shown in Fig. 2d and provides better visual correspondence with ground truth image in Fig. 2b. We have estimated from Figs. 2c and 2d that $\mu(\Psi(\mathbf{A}))\approx 0.9760$, which again confirms capability of the EKM projection to decrease similarity between data points. The FVS method estimated dictionary from 6000 randomly selected pixels. The color coded result of LSC algorithm [16] is shown in Fig. 2e. The best result, shown in Fig. 2e, is obtained for hand tuning the number of neighbors from 2 to 5. As shown in [1], state-of-the-art image segmentation methods such as *k-means* clustering in the CIE $L^*a^*b^*$ color space [19], is also shown in Fig. 2f. The FELRSC algorithm finished much faster than other compared algorithms. The computation time of the FELRSC algorithm was 1.63 seconds compared to 250 seconds needed for the EKM-NMF_L0 and 199 seconds for the LSC algorithm.

Table 1. Clustering accuracy ± standard deviation (%) evaluated after 10 runs on synthetic dataset: *original data, ** Gaussian kernel based EKM-projected data with variance according to (10). SNR is in dB.

SNR	FELRSC	LSC	LRR*	LRR**	NSN-GSR*	NSN-GSR**
70	100	82.3 ±0.2	100	100	100	100
40	100	82.3 ±0.2	56.1 ±0.5	100	100	100
20	100	82.2 ±0	41.3 ±0.6	77 ±10.6	72.1 ±13.4	90 ±0.9
10	95.7 ±0.9	82.3 ±0	37.8 ±0.5	45.5 ±4.1	67.9 ±1.6	88.7 ±3.1

We have also applied the proposed FELRSC algorithm to segment microscopic image of unstained specimen of human liver with metastasis of colon cancer, Fig. 3a. The image is of the same size as the one shown in Fig. 2a. Thus, equivalent comments apply to computational intractability of spectral clustering-based SC algorithms. We have estimated from Fig. 3a that $\mu(\mathbf{A})=0.9998$ [1]. The main tissues presented in the image are metastasis of colon cancer, hepatocytes and border area between the tumor and liver tissue [1]. The ground truth image related to image of unstained specimen shown in Fig. 3a is obtained by immunohistochemical staining subsequent sections with Hep Par (Figure 3b) and CDX2 (Figure 3c). In Hep Par, the ground truth image hepatocytes are colored brown, whereas the metastatic cells of colon cancer and the inflammatory cells are blue. In CDX2 the ground truth images metastasis of colon cancer is colored completely or partially brown, whereas the hepatocytes and inflammatory cells are colored blue. The color coded result obtained proposed FELRSC algorithm is shown in Fig. 3e and by the EKM-NMF_L0 algorithm in Fig. 3f. Thereby, metastasis of colon cancer is colored in blue, border area between the tumor and liver tissue in red and hepatocytes in green. Color coded results obtained by the FELRSC algorithm

provides more convincing visual correspondence with ground truth image in Fig. 3b and 3c for both metastasis of colon cancer and hepatocytes. Computation times of the algorithms are the same as for Fig. 2.

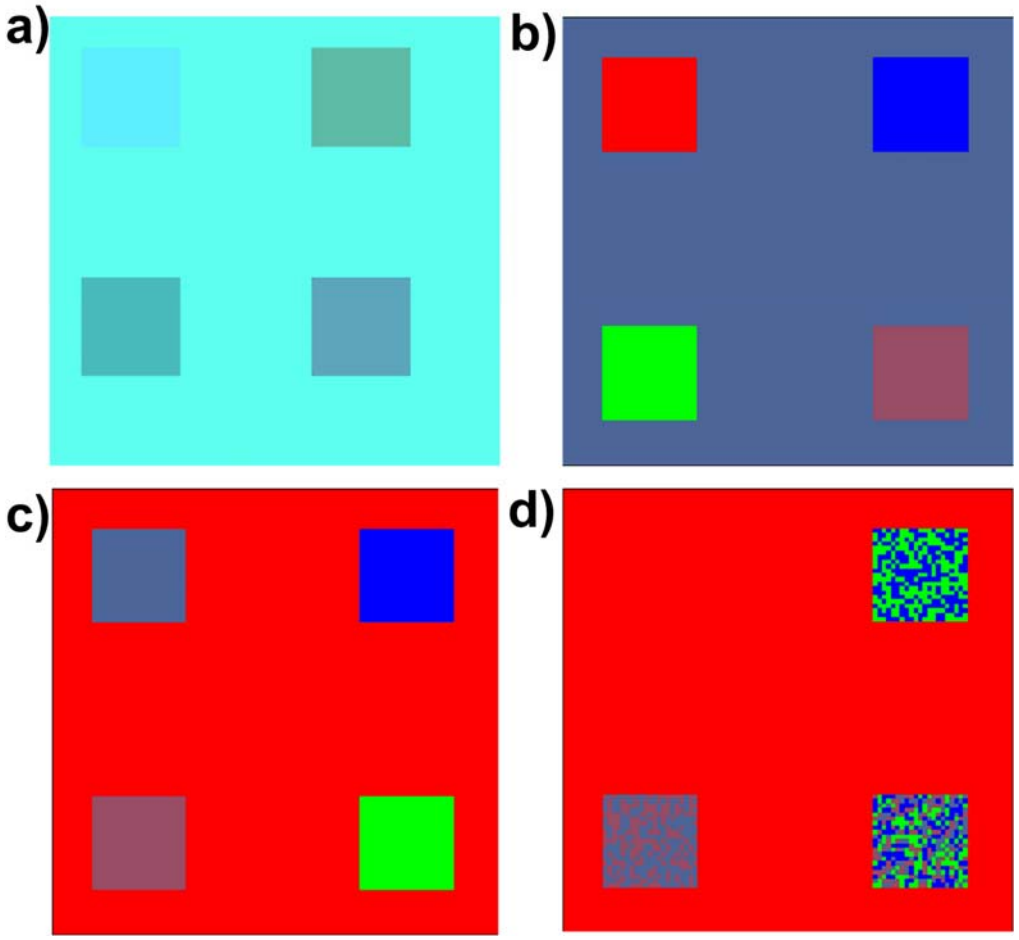

Figure 1. (a) synthetic image with SNR=20dB in red-green-blue color space. Color coded: (b) ground truth; (c) the FELRSC result; d) the NSN-GSR** result.

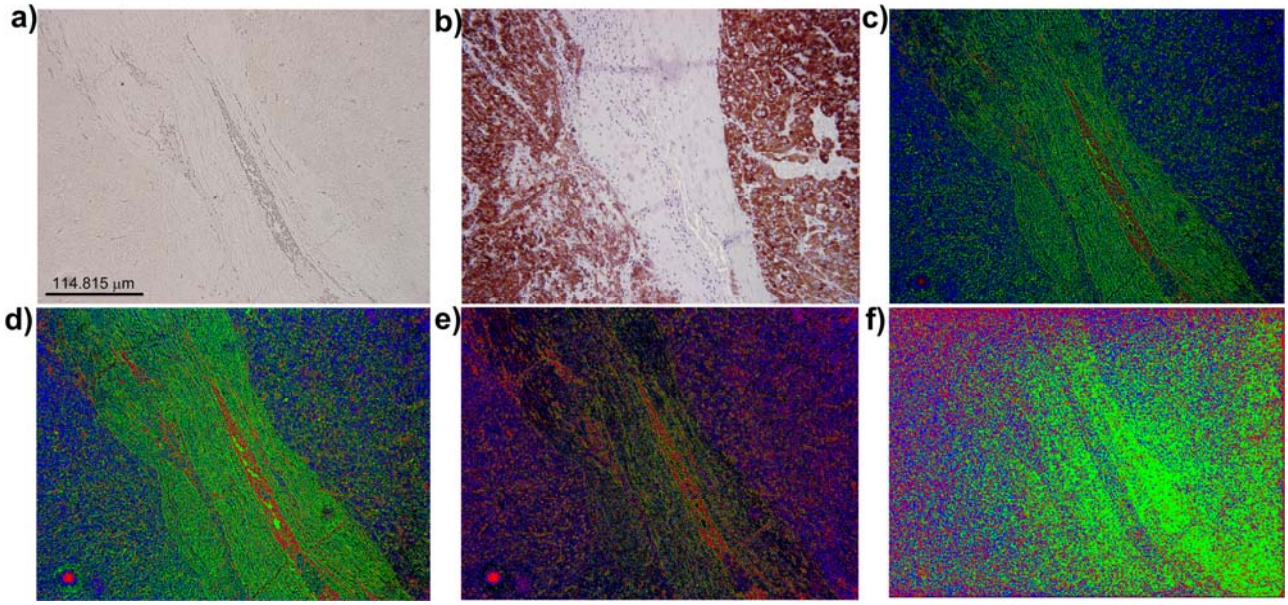

Figure 2. Image of: a) unstained specimen of human liver with hepatocellular carcinoma. Size of the image $N \approx 1.44 \times 10^6$ pixels; b) HepPar stained specimen in subsequent slide. Color coded results of: c) the EKM-NMF_L0 [1]; d) the proposed FELRSC algorithm; e) the LSC [16]; f) *k-means* in CIE La^*b^* color space [19]. Blue: HCC; green: tumor fibrotic capsule; red: blood vessel.

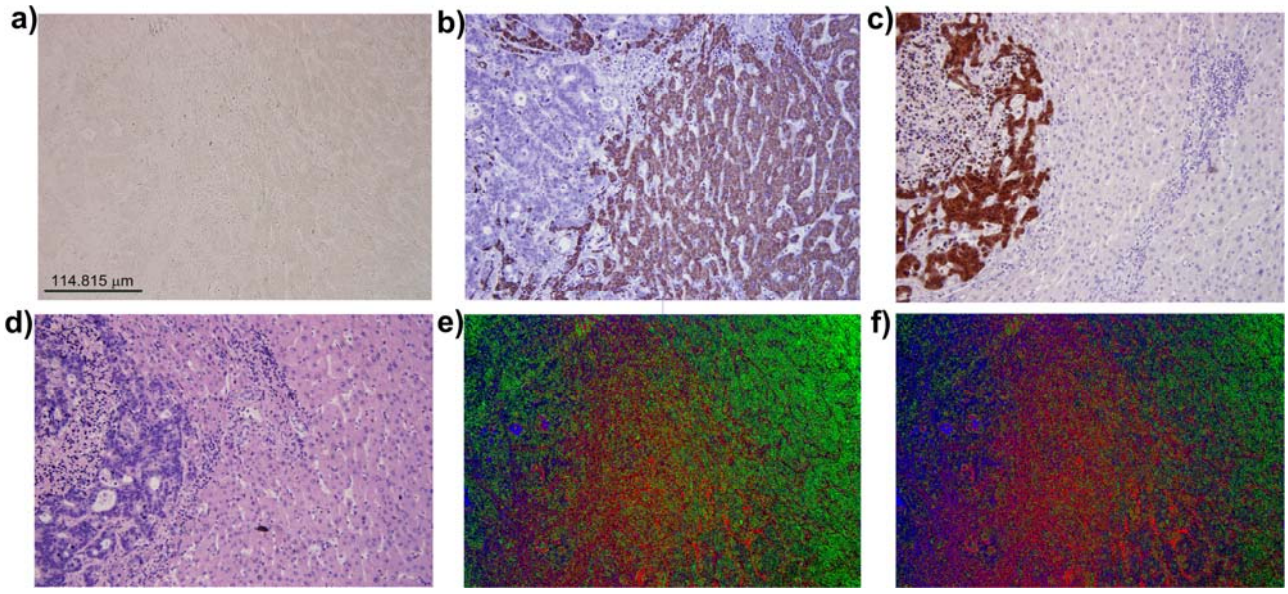

Figure 3. Image of: a) unstained specimen of human liver with metastasis of colon cancer. Size of the image $N \approx 1.44 \times 10^6$ pixels; b) HepPar stained specimen in subsequent slide; c) CDX2 stained specimen in subsequent slide; d) Color image of the specimen a) stained by H&E. Color coded results of: e) the proposed FELRSC algorithm; f) EKM-NMF_L0 [1]. Blue: metastasis of colon cancer; green: hepatocytes; red: border area between tumor and liver tissue.

ACKNOWLEDGEMENTS

This work was supported the through Grant "Structured Decompositions of Empirical Data for Computatioally-Assisted Diagnoses of Disease" funded by Croatia Science Foundation and through bilateral Chinese-Croatian grant "Decomposition and Contrast Enhancement of PET/CT and OCT images by means of nonlinear sparse component analysis".

REFERENCES

[1] Kopriva, I., Popović Hadžija, M., Hadžija, M., Aralica, G., "Unsupervised segmentation of low-contrast multichannel images: discrimination of tissue components in microscopic image of unstained specimen," Scientific Reports 5, article no. 11576 (2015).
[2] Vidal, R., "Subspace clustering," IEEE Sig. Proc. Mag. 28 (2), 52–68 (2011).
[3] Elhamifar, E., Vidal, R., "Sparse Subspace Clustering: Algorithm, Theory, and Applications," IEEE Trans. Patt. Anal. Mach. Intel. 35 (1), 2765-2781 (2013).
[4] Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y., "Robust recovery of subspace structures by low-rank representation," IEEE Trans. Patt. Anal. Mach. Intel. 35 (1), 171–184 (2013).
[5] Ng, A. Y., Jordan, M. I., Weiss, Y., "On spectral clustering: Analysis and an algorithm," Proc. Neural Information Processing Systems 14, 849–856 (2011).
[6] Park, D., Caramanis, C., Sanghavi, S., "Greedy Subspace Clustering," Proc. Neural Information Processing Systems 27, 2753-2761 (2014).
[7] Heckel, R., Bölcskei, H., "Robust Subspace Clustering via Thresholding," IEEE Trans. Inf. Theory 61, 6320-6342 (2015).
[8] Wang, Y., Tang, Y. Y., Li, L., "Minimum Error Entropy Based Sparse Representation for Robust Subspace Clustering," IEEE Trans. Sig. Proc. 63 (15), 4010-4021 (2015).
[9] Vidal, R., Favaro, P., "Low rank subspace clustering (LRSC)," Patt. Rec. Lett. 43, 47-61 (2014).
[10] http://fulir.irb.hr/id/eprint/2408
[11] Shawe-Taylor, J. and Cristianini, N., [Kernel Methods for Pattern Analysis], Cambridge Univ. Press (2004).
[12] Kopriva, I., Jerić, I., Brkljačić, L., "Nonlinear Mixture-wise Expansion Approach to Underdetermined Blind Separation of Nonnegative Dependent Sources," J. of Chemometrics 27 (7-8), 189-197 (2013).
[13] Donoho, D. L., Huo, X., "Uncertainty principles and ideal atomic decomposition," IEEE Trans. Inf. Theory 47, 2845–2862 (2001).
[14] Vidal, R., Favaro, P., "Low rank subspace clustering (LRSC)," Patt. Rec. Lett. 43, 47-61 (2014).
[15] http://dhpark22.github.io/greedysc.html
[16] Cai, D., Chen, X., "Large Scale Spectral Clustering Via Landmark-Based Sparse Representation," IEEE Trans. on Cybernetics 45(6), 1669-1680 (2015).
[17] http://www.cad.zju.edu.cn/home/dengcai/Data/Clustering.html
[18] Baudat, G., Anouar, F., "Kernel-based methods and function approximation," Proc. Int. Joint Conference on Neural Networks, Washington, DC, 1244-1249 (2001).
[19] Chitade, A. Z., Katiyar, S. K., "Colour Based Image Segmentation Using K-Means Clustering," Int. J. Eng. Sci. Tech. 2, 5319–5325 (2010).